L2/07-413

Proposal to Encode Oriya Fraction Signs in ISO/IEC 10646

Anshuman Pandey
University of Michigan
Ann Arbor, Michigan, U.S.A.
pandey@umich.edu

December 4, 2007

Contents

List of Figures

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]

Please fill all the sections A, B and C below. Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest Roadmaps.

A. Administrative

1. Title: **Proposal to Encode Oriya Fraction Signs in ISO/IEC 10646**
2. Requester's name: **Anshuman Pandey (pandey@umich.edu)**
3. Requester type (Member Body/Liaison/Individual contribution): **Individual contribution**
4. Submission date: **December 4, 2007**
5. Requester's reference (if applicable): **N/A**
6. Choose one of the following:
   (a) This is a complete proposal: **Yes**
   (b) or, More information will be provided later: **No**

B. Technical - General

1. Choose one of the following:
   (a) This proposal is for a new script (set of characters): **No**
      i. Proposed name of script: **N/A**
   (b) The proposal is for addition of character(s) to an existing block: **Yes**
      i. Name of the existing block: **Oriya**
2. Number of characters in proposal: **6**
3. Proposed category: **A - Contemporary**
4. Is a repertoire including character names provided?: **Yes**
   (a) If Yes, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?: **Yes**
   (b) Are the character shapes attached in a legible form suitable for review?: **Yes**
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?: **Anshuman Pandey**; **True Type**
   (a) If available now, identify source(s) for the font and indicate the tools used: **The font contains normalized forms of the Oriya fraction signs as found in printed documents. It was drawn by Anshuman Pandey using Metafont and converted to True Type format using FontForge.**
6. References:
   (a) Are references (to other character sets, dictionaries, descriptive texts etc.) provided?: **Yes**
   (b) Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?: **Yes**
7. Special encoding issues:
   (a) Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? **No**
8. Additional Information: Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard. **Character properties, numeric information, and currency information are included.**

[1] Form number: N3102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?: **No**
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? **No**
   (a) If Yes, with whom?: **N/A**
      i. If Yes, available relevant documents: **N/A**
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **Yes**
   (a) Reference: **The signs were used by the Oriya-speaking community.**
4. The context of use for the proposed characters (type of use; common or rare): **Common**
   (a) Reference: **The signs were used for numeric notation in the Oriya script.**
5. Are the proposed characters in current use by the user community?: **No**
   (a) If Yes, where? Reference: **The signs are not used at present.**
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?: **Yes**
   (a) If Yes, is a rationale provided?: **The signs belong to the Oriya script, which is encoded in the BMP. There is sufficient space in the Oriya block for the inclusion of the signs.**
      i. If Yes, reference: **N/A**
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **Yes. The characters constitute a specialized set.**
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? **No**
   (a) If Yes, is a rationale for its inclusion provided?: **N/A**
      i. If Yes, reference: **N/A**
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **No**
   (a) If Yes, is a rationale provided?: **N/A**
      i. If Yes, reference: **N/A**
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? **Yes**
    (a) If Yes, is a rationale for its inclusion provided? **Yes**
       i. If Yes, reference: **See text of proposal.**
11. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? **No**
    (a) If Yes, is a rationale for such use provided? **N/A**
       i. If Yes, reference: **N/A**
    (b) Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **No**
       i. If Yes, reference: **N/A**
12. Does the proposal contain characters with any special properties such as control function or similar semantics? **No**
    (a) If Yes, describe in detail (include attachment if necessary): **N/A**
13. Does the proposal contain any Ideographic compatibility character(s)? **No**
    (a) If Yes, is the equivalent corresponding unified ideographic character(s) identified? **N/A**
       i. If Yes, reference: **N/A**

1 Introduction

This is a proposal to encode Oriya fraction signs as part of the Oriya script in the Basic Multilingual Plane (BMP) of the Universal Character Set (UCS) (ISO/IEC 10646).

The encoding of Oriya fraction signs is necessary for the full and accurate reproduction of historical numeric notation in the Oriya script and for such notation in digital media.

2 Characters Proposed

Characters The characters proposed for the six Oriya fractions are:

- ORIYA FRACTION ONE QUARTER
- ORIYA FRACTION ONE HALF
- ORIYA FRACTION THREE QUARTERS
- ORIYA FRACTION ONE SIXTEENTH
- ORIYA FRACTION ONE EIGHTH
- ORIYA FRACTION THREE SIXTEENTHS

Basis for Character Shapes The forms of the proposed characters are based on normalized forms of the respective characters found in printed documents (see Figure 1 and Figure 2).

3 Technical Features

Name The names of the Oriya fraction signs are based on the fractional values they represent.

Allocation It is recommended that the Oriya fractions be encoded at the range U+0B72..U+0B77.

Properties The fraction signs belong to the Unicode general category "Number, Other" (No). The appropriate numeric value is assigned to each sign: "1/16" for ORIYA FRACTION ONE SIXTEENTH, "1/2" for ORIYA FRACTION ONE HALF, etc. The fraction signs are always written left-to-right and have bidirectional values of "Left-to-Right" (L).

```
0B72;ORIYA FRACTION ONE QUARTER;No;0;L;;;;1/4;N;;;;;
0B73;ORIYA FRACTION ONE HALF;No;0;L;;;;1/2;N;;;;;
0B74;ORIYA FRACTION THREE QUARTERS;No;0;L;;;;3/4;N;;;;;
0B75;ORIYA FRACTION ONE SIXTEENTH;No;0;L;;;;1/16;N;;;;;
0B76;ORIYA FRACTION ONE EIGHTH;No;0;L;;;;1/8;N;;;;;
0B77;ORIYA FRACTION THREE SIXTEENTHS;No;0;L;;;;3/16;N;;;;;
```

4 Overview of the Signs

Description The fraction signs were used for numeric notation in the Oriya script. The signs represent fraction values of a base-16 system. They appear in both written and printed materials. The fraction signs are no longer used in Orissa. The change of the currency system to a decimal system in 1957 and the adoption of the metric system for weights and measures in 1958 rendered obsolete the number forms used for the base-16 system.

Orthography The fractions of the base-16 system are written using the six proposed characters. All fractions can be created from these characters through an additive process. The pattern is as follows:

1/16	୵	5/16	୷	9/16	୸	13/16	୹
1/8	୶	3/8	୷୶	5/8	୸୶	7/8	୹୶
3/16	୷	7/16	୷୷	11/16	୸୷	15/16	୹୷
1/4	୳	1/2	୸	3/4	୹	1	୰

The additive principle also underlies the Bengali method of writing fractions. There is no distinct sign for 1/2 in Bengali, instead the value is represented by writing ৷ U+09F7 BENGALI CURRENCY NUMERATOR FOUR twice. The Bengali pattern is as follows:

1/16	৵	5/16	৷৵	9/16	৷৷৵	13/16	৸৵
1/8	৶	3/8	৷৶	5/8	৷৷৶	7/8	৸৶
3/16	৷	7/16	৷৷	11/16	৷৷৷	15/16	৸৷
1/4	৷	1/2	৷৷	3/4	৸	1	৯

5 Relationship to Other Signs

Three of the Oriya fraction signs are similar to the currency numerators of the Bengali script. The ୳ ORIYA FRACTION ONE QUARTER is graphically similar to ৷ U+09F7 BENGALI CURRENCY NUMERATOR FOUR (as well as to ৷ U+A830 NORTH INDIC FRACTION ONE QUARTER). The ୵ ORIYA FRACTION ONE SIXTEENTH is semantically and graphically similar to ৵ U+09F4 BENGALI CURRENCY NUMERATOR ONE, as is ୶ ORIYA FRACTION ONE EIGHTH to ৶ U+09F5 BENGALI CURRENCY NUMERATOR TWO. The signs ୸ ORIYA FRACTION ONE HALF, ୹ ORIYA FRACTION THREE QUARTERS, and ୷ ORIYA FRACTION THREE SIXTEENTHS are distinct from their Bengali counterparts.

The following compares the base fraction signs of the Oriya, Bengali, and the north Indic numeric notation systems:

	୵	ORIYA FRACTION ONE SIXTEENTH
1/16	৵	U+09F4 BENGALI CURRENCY NUMERATOR ONE
	𑁓	U+A833 NORTH INDIC FRACTION ONE SIXTEENTH
	୶	ORIYA FRACTION ONE EIGHTH
1/8	৶	U+09F5 BENGALI CURRENCY NUMERATOR TWO
	𑁔	U+A834 NORTH INDIC FRACTION ONE EIGHTH
	୷	ORIYA FRACTION THREE SIXTEENTHS
3/16	৷	U+09F6 BENGALI CURRENCY NUMERATOR THREE
	𑁕	U+A835 NORTH INDIC FRACTION THREE SIXTEENTHS
	୳	ORIYA FRACTION ONE QUARTER
1/4	৷	U+09F7 BENGALI CURRENCY NUMERATOR FOUR
	𑁒	U+A830 NORTH INDIC FRACTION ONE QUARTER
	୸	ORIYA FRACTION ONE HALF
1/2	৷৷	U+09F7 BENGALI CURRENCY NUMERATOR FOUR (written twice)
	𑁑	U+A831 NORTH INDIC FRACTION ONE HALF
	୹	ORIYA FRACTION THREE QUARTERS
3/4	৸	U+09F8 BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR
	𑁖	U+A832 NORTH INDIC FRACTION THREE QUARTERS

Although the Oriya and Bengali fraction signs represent values of the same base-16 system, the manner in which the Bengali currency numerators were defined in the UCS necessitates that the Oriya characters be defined differently than their Bengali counterparts. The two issues are names and properties.

Despite the semantic and graphical similarity between some Oriya and Bengali signs, the descriptor "currency numerator" used in the names of Bengali signs is inappropriate for characters used for general numeric notation, not solely for currency. The description "fraction" more adequately reflects the larger application of the proposed characters. The names of the Oriya fraction signs are based on the model of the north Indic fraction signs accepted for encoding in the Common Indic Number Forms block.

Second, the numeric values of the Bengali currency numerators are integers:

```
09F4;BENGALI CURRENCY NUMERATOR ONE;No;0;L;;;;1;N;;;;;
09F5;BENGALI CURRENCY NUMERATOR TWO;No;0;L;;;;2;N;;;;;
09F6;BENGALI CURRENCY NUMERATOR THREE;No;0;L;;;;3;N;;;;;
09F7;BENGALI CURRENCY NUMERATOR FOUR;No;0;L;;;;4;N;;;;;
09F8;BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR;No;0;L;;;;;N;;;;;
09F9;BENGALI CURRENCY DENOMINATOR SIXTEEN;No;0;L;;;;16;N;;;;;
```

These numeral values are not appropriate for characters that fundamentally represent fractions. For example, ୵ ORIYA FRACTION ONE EIGHTH and ৵ U+09F5 BENGALI CURRENCY NUMERATOR TWO are used in the respective scripts to represent ⅛; however, while the proposed ORIYA FRACTION ONE EIGHTH is assigned the numeric value ⅛, the U+09F5 BENGALI CURRENCY NUMERATOR TWO is assigned the numeric value 2.

For this reason, the properties assigned to the north Indic fraction signs are more appropriate for the Oriya fraction signs:

```
A830;NORTH INDIC FRACTION ONE QUARTER;No;0;L;;;;1/4;N;;;;;
A831;NORTH INDIC FRACTION ONE HALF;No;0;L;;;;1/2;N;;;;;
A832;NORTH INDIC FRACTION THREE QUARTERS;No;0;L;;;;3/4;N;;;;;
A833;NORTH INDIC FRACTION ONE SIXTEENTH;No;0;L;;;;1/16;N;;;;;
A834;NORTH INDIC FRACTION ONE EIGHTH;No;0;L;;;;1/8;N;;;;;
A835;NORTH INDIC FRACTION THREE SIXTEENTHS;No;0;L;;;;3/16;N;;;;;
```

6 References

Pihan, Antoine Paulin. 1860. *Exposé des signes de numération usités chez les peuples orientaux anciens et modernes*. Paris: L'imprimerie impériale.

Young, A. H. 1935. *First Lessons in Oriya*. Revised by B. Das. Cuttack: Orissa Mission Press.

100 NUMÉRATION OURIYA.

CHIFFRES.	VALEURS.	NOMS DE NOMBRE.	CHIFFRES.	VALEURS.	NOMS DE NOMBRE.
୮୫	85	*pañtchâçi.*	୨୩	93	*téyânabé.*
୮୬	86	*tchhayâçi.*	୨୪	94	*tchourânabé.*
୮୭	87	*satâçi.*	୨୫	95	*pañtchânabé.*
୮୮	88	*aṭhâaçi.*	୨୬	96	*tchhayânabé.*
୮୨	89	*ounnabé.*	୨୭	97	*satânabé.*
୨୦	90	*nabé.*	୨୮	98	*aṭhânabé.*
୨୧	91	*ékânabé.*	୨୨	99	*ounçayé.*
୨୨	92	*byânabé.*	୧୦୦	100	*çayé.*

AUTRES NOMBRES PLUS ÉLEVÉS.

sahasr, hadjâr.. mille;
ayout..... dix mille;
lakch..... cent mille;
niyouté..... 1 suivi de six zéros;
koṭié..... 1 ..... sept zéros;
arboud..... 1 ..... huit zéros;
mahârboud.... 1 ..... neuf zéros;
padma..... 1 ..... dix zéros;
mahâpaḍma... 1 ..... onze zéros;
kharb..... 1 ..... douze zéros;
mahâkharb.. 1 suivi de treize zéros;
çañkh..... 1 ..... quatorze zéros;
mahâçañkh.. 1 ..... quinze zéros;
hâhâ..... 1 ..... seize zéros;
mahâhâhâ... 1 ..... dix-sept zéros;
dhoûl..... 1 ..... dix-huit zéros;
mahâdhoûl... 1 ..... dix-neuf zéros;
akchôhiṇi.... 1 ..... vingt zéros;
mahâkchôhiṇi. 1 ..... vingt et un zéros.

FRACTIONS.

୧ ¼ *pâé* ou *pâddoulâ*;
୪ ½ *adh, ardh, arek*;
୩ ¾ *paounou* ou *tinpâé*;
୧୪ 1 ½ *doḍh*;
୨୪ 2 ½ *aḍhâi*.

Figure 1: Method of writing fractions in the Oriya script (from Pihan, 1860: 100).

Fractions.

The leading principle of Oriya arithmetic, to divide by *four* rather than any other number, pervades also the system of fractions.

୲ means one quarter of the unit, rupee, or maund.
୳ „ two quarters.
୴ „ three quarters.
୵ „ one fourth of a quarter, or one sixteenth of the unit.
୶ „ two sixteenths.
୷ „ three sixteenths.

Figure 2: Method of writing fractions in the Oriya script (from Young, 1935: 203).